

Dating in DAMBE with internal-calibration

DAMBE does dating with either strict clock or relaxed clock. You choose to use strict or relaxed clock by unchecking or checking the 'Relaxed clock' checkbox (which is checked by default)

The input is

1. A bifurcating file (if the tree is trifurcating, just drag one child of the root to another to make the tree bifurcating).
2. Aligned sequence file or distance matrix file
3. One or more calibration points (when the tree is displayed,

The OTU names in the tree file needs to be the same as those in the sequence or distance files.

Here are the file pairs that are included in DAMBE installation under directory C:\Program Files\DAMBE (for 32-bit machine) or C:\Program Files (x86)\DAMBE:

1. Aligned sequence files (.NEX) and their associated topologies (.dnd)
 - a. **mtDNAPri3.NEX** and **mtDNAPri.DND**: The nexus file contains third codon positions of the 12 protein-coding genes collinear with the L-strand of vertebrate mitochondrial genome for seven great ape species. The sequences are used in Rannala and Yang (2007), with two reasonably good calibration points.
 - b. **PrimateOnly.NEX** and **PrimateOnlyBEAST.dnd**: The nexus file contains 26 primate species with three reasonably good calibration points. They are used in Yang and Yoder (2003).
 - c. **MouseLemurs3.NEX** and **MouseLemurs.DND**: Include more sequences than those in b
 - d. **VertCOLFAS** and **vertCOIRooted.dnd**
2. Distance files (.dis) and associated topologies (.dnd):
 - a. **KumarFig2.dis** and **KumarFig2.dnd**: taken from an unpublished manuscript by S. Kumar and colleagues, with extreme violation of the clock hypothesis.
 - b. **mtDNAPri3F84SE.dis** and **mtDNAPri.DND**:
 - c. **mtDNAPri3F84SEMult.dis** (which contains 100 distance matrices) and **mtDNAPri.DND**

Dating with aligned sequences: We will first perform a simple dating, and then estimate variability of the estimates.

Simple dating: Click the DAMBE icon to start DAMBE. Click 'File|Open standard sequence file'. In the dialog box Fig. 12-1, click the 'File of type' dropdown box to select 'PAUP/NEXUS' format. All files with file types '.nex', '.pau', '.paup' will be displayed (Fig. 12-1). Double-click mtDNAPri3.NEX to open it (or just single-click to highlight it and then click the 'Open' button). If you want to try some sequence files in other format, just click the 'File of type' dropdown box and choose your file type. If you are not sure of your sequence format, choose 'Unknown' to let DAMBE decide.

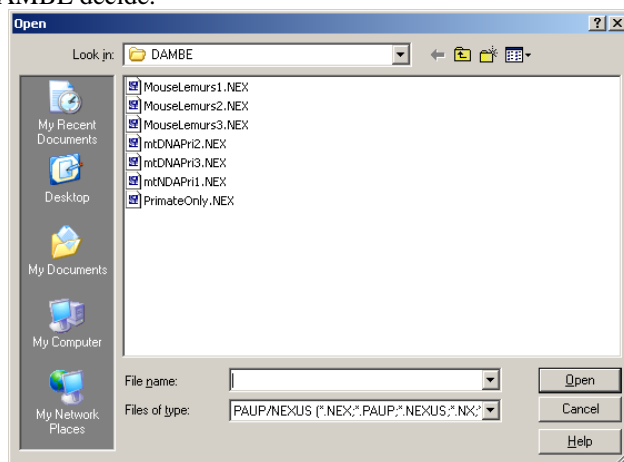


Fig. 12-1. File open dialog box in DAMBE.

You will be asked for sequence information (Fig. 12-2). Different types of sequences are associated with different analytical methods. ‘Protein-coding Nuc. Seq.’ means protein-coding nucleotide sequences that can be translated into amino acids. In our case, although the third codon positions are from a protein-coding gene, the sequence as a whole is not protein-coding. So you should keep the default ‘Non-protein Seq.’ and click the ‘Go!’ button.

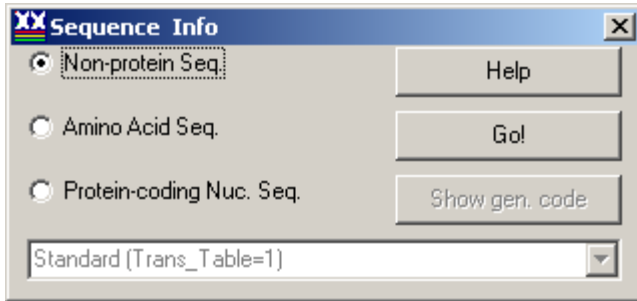


Fig. 12-2. Dialog for sequence information.

The aligned sequences will be displayed. Now click ‘Phylogenetics|Distance-based dating’. The next dialog box (Fig. 12-3) is for you to specify the rooted tree and the genetic distance to generate from the sequences. In earlier versions of DAMBE, the ‘Relaxed clock’ is grayed (not enabled). It is enabled (checked) by default since DAMBE version 7.035. For lineages that evolve at dramatically different rates, you should leave ‘Relaxed clock’ checked. It adds a little bit of computation, but the dating results should be more satisfactory.

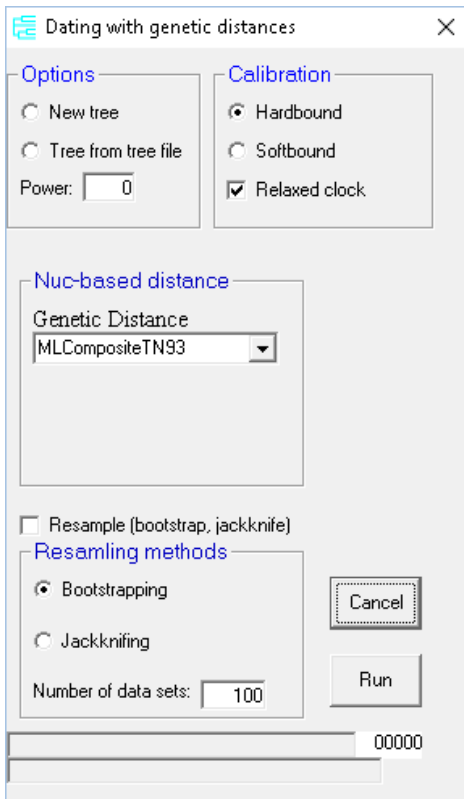


Fig. 12-3. Dialog box to specify options in distance-based dating.

You can input a rooted tree in two ways. We will use a tree that has already been generated and saved in the file mtDNAPri.dnd. Click ‘Tree from tree file’ at the dialog box shown in Fig. 12-3, browse to DAMBE’s installation directory (e.g., C:\Program Files\DAMBE), and open the mtDNAPri.dnd file. The rooted tree will be displayed in the tree panel (Fig. 12-4). Note that it is a rooted tree with all nodes bifurcating.

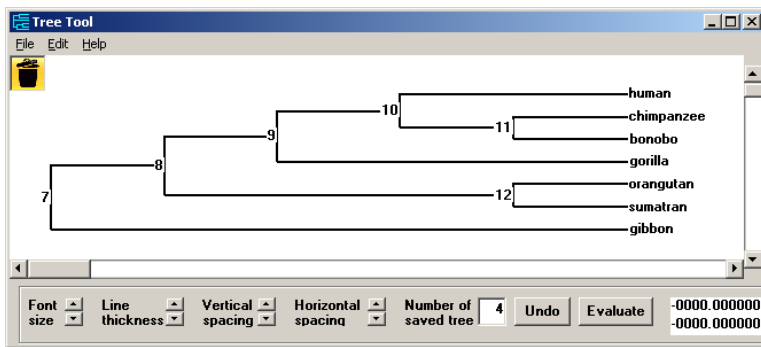


Fig. 12-4. Tree panel for displaying or modifying topologies.

(If you want to use your own tree, make sure that the OTU names in the tree match those in the sequences. If you do not have a tree but know what the topology should look like, you can click the ‘New tree’ option button. This will bring up a tree panel with a star tree with OTUs being the names of the sequences that have been read into DAMBE. You can then drag species or nodes to each other to form a rooted tree.)

You now need to specify calibration points. Note that all the internal nodes are numbered. Two calibration points have been used for dating this set of sequences (Rannala and Yang 2007), one for the common ancestor of all three great apes and human (node 8), and the other for the common ancestor of human and chimpanzee (node 10). Right-click (Mac users: Enable the secondary click by go to ‘System Preferences|Keyboard & Mouse’ and set the right button to ‘Secondary Button’. The Mac default is to have both the left and right buttons set to the same ‘Primary Button’. You may also set the tiny middle button, which defaults to ‘dashboard’, to the secondary button) the node numbered ‘8’ (including human, two chimp species, gorilla and two orangutan species), and enter 14 (for 14 million years ago) in the ensuing dialog box. Right-click the node numbered ‘10’ (including human, and two chimp species) and enter 7 (for 7 million years ago). Note that the calibration nodes are now coloured in red. (You can click the tree in many different ways. Different clicks are often associated with advanced dating functions. I have not finalized the interface. So click only as instructed.)

One should be aware of possible inconsistency in calibration points. Take Fig. 12-4 for example, if the common ancestor for human and chimpanzee (node 10) was dated to be 14 million years old, whereas the common ancestor for human, chimp, gorilla and orangutan (node 8) was dated to be less than 14 million years old, then we have inconsistent calibration points. The dating algorithm will fail with inconsistent calibration points.

Click ‘File|Export and exit’ to quit the tree panel. If you have forgotten to set the calibration points, DAMBE will give you an opportunity to set them. If you have followed the instructions, the tree panel will close and the tree-selection option in the dialog box (Fig. 12-3) will now be disabled because the selection is successful.

We have more options to set in the dialog box shown in Fig. 12-3. If you choose the ‘Hardbound’ option, then the calibration times will be fixed. If you choose the ‘Softbound’ option, then the calibration points will be changed to minimize the residual sum of squares when multiple calibration points are available. The ‘Softbound’ option will have no effect when there is only a single calibration point.

I suggest that you choose the ‘Softbound’ option. This is not only because the ‘Softbound’ option will lead to smaller residual sum of squares, but also because fossil dating is often inaccurate. The ‘softbound’ option allows the use of observed data to revise the calibration points. In a sense, the original calibration points may be taken as a prior estimate, which is revised to give a posterior estimate by incorporating information from the observed data.

In the genetic distance, the default is ‘MLCompositeF84’. This is one of the simultaneously estimated (SE) distances for the F84 substitution model (Tamura, et al. 2004; Xia 2009). It is crucial to use SE distances, especially when using the third codon position, because most of the independent estimation method will generate inapplicable cases due to high sequence divergence.

Two such SE approaches have been implemented in DAMBE, one with the maximum likelihood method and the other with the LS method (Xia 2009). These distances are implemented for the F84 and TN93 substitution models in DAMBE and designated ‘MLCompositeF84’, ‘MLCompositeTN93’, ‘LSCompositeF84’, and ‘LSCompositeTN93’ (where ML stands for maximum likelihood and LS stands for least-squares). For the F84 model, a global transition/transversion ratio will be estimated and used for all pairwise distance calculations. For the TN93 model,

there are two transition/transversion ratios (κ_1 and κ_2) which are used for all pairwise distance calculations. This approach essentially eliminates the inapplicable cases, and produces much more robust distance estimates.

In this first try, we will not do resampling. So leave the resampling checkbox cleared and click the 'Run' button. The output is displayed in two parts. The first is a tree with divergence time estimates (Fig. 12-5). Note that the two calibration points are not 14 and 7 as we have inputted, but are 14.75 and 5.5, respectively. Had you chosen "Hardbound" instead of 'Softbound' in the dialog box in Fig. 12-3, the two calibration nodes will show exactly as the input values, i.e., 14 and 7, respectively. The option 'Softbound' allows DAMBE to fine-tune the time of the calibration nodes by minimizing the sum of residual squares.

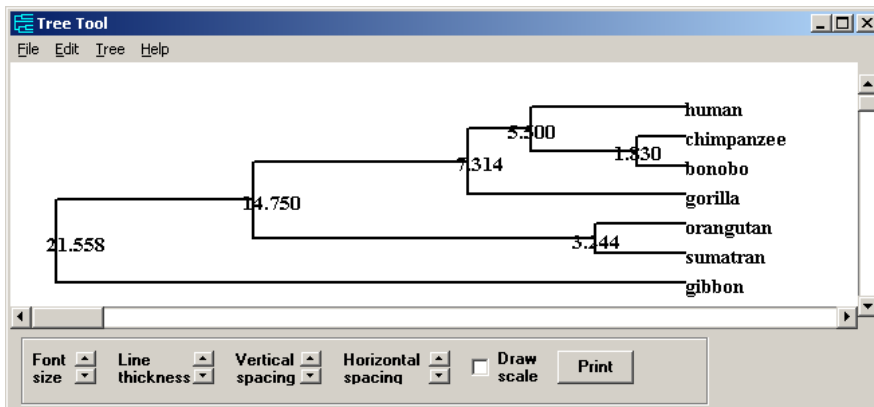


Fig. 12-5. Tree with estimated divergence time.

The tree can be manipulated in many ways and can be exported in low-resolution bitmapped .bmp format or the high-resolution metafile format either for publication or PowerPoint presentation.

The second part is the text output, which includes the distance matrix, the estimated parameters of the substitution model such as the transition/transversion ratio, log-likelihood for the distance estimation, and details of the dating input and results such as the evolutionary rate, the tree with dates and the residual sum of squares (RSS). RSS will be larger if you had chosen the 'Hardbound' option.

You are encouraged to explore with various combinations of 'Hardbound' or 'Softbound', 'Relaxed clock' enabled or disabled, etc.

Estimating variability of the estimated divergence time: Now repeat the procedure until you get to the dialog box shown in Fig. 12-3. Read in the mtDNAPri.DND tree and set the two calibration times as before. Now click the 'Resample (bootstrap, jackknife)' check box. The default resampling protocol is 'Bootstrap' and the default number of resampled data sets is 100. The higher the number, the better the estimates are. For practical research, the number of resampled data sets should be at least 500. Click the 'Run' button.

After DAMBE has bootstrapped the specified number of times, dating results are again presented in two parts: a tree (Fig. 12-6) with divergence times as well as text output. The tree differs from that in Fig. 12-5 in that the standard deviation of the estimated divergence time is also presented.

It is important to keep in mind that the divergence time will be slightly different when you repeat the bootstrapping procedure because bootstrapping randomly samples sites to reconstitute a new set of sequences.

You should now use the second set of sequences (PrimateOnly.NEX) and the associated tree (PrimateOnlyBEAST.NEX) to see how well the distance-based dating method work with a larger data set. Note that this set of sequences contains unresolved nucleotides and DAMBE will ask you how to handle them. You may just leave the default, i.e., '3 leaving them as they are'. This will result in ambiguous codes treated as follows. If it is an R (standing for A or G) and if the nucleotide frequencies of A and G are, say, 0.4 and 0.1, then R will be treated as A with a probability of 0.8 and as G with a probability of 0.2. Of course, if A and G are equal, then R will be treated as 0.5 A and 0.5 G. The same rule applies to other ambiguous codes.

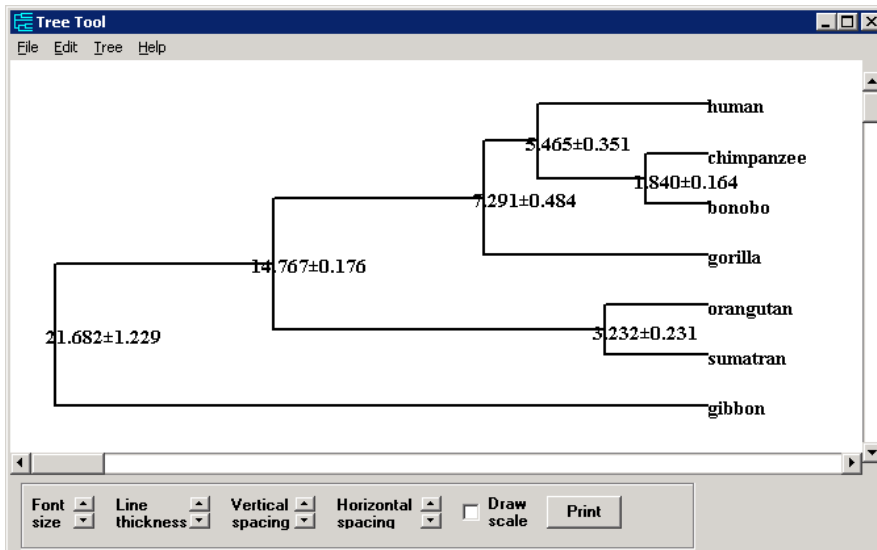


Fig. 12-6. Tree with the estimated divergence time and the associated standard deviation.

You can set three calibration points (Yang and Yoder 2003) when dating this set of sequences, one at the root with 77 million years, one internal node including Homo, Pan, Gorilla, Pongo, and Macaca with 35 million years, and the third node including Homo, Pan, and Gorilla with 10 million years. Continue until you get your dating results.

Dating with distance matrices

When only a single distance matrix is used, then the dating output cannot have any estimate of the variability of the estimated divergence time. With multiple distance matrices, however, each matrix will generate one set of estimated divergence time and a measure of the variability of the divergence times can then be estimated.

Click 'File|Open other molecular data file', choose 'Distance matrix file', and click the 'OK' button. In the 'Open file' dialog box, choose the 'mtDNAPri3F84SE.dis' file and click the 'Open' button. This file contains a single distance matrix and will generate one set of divergence time estimates. In the ensuing dialog box (Fig. 12-7), change the default from 2 (Testing the molecular clock hypothesis) to 3 (Dating). Click 'OK'.

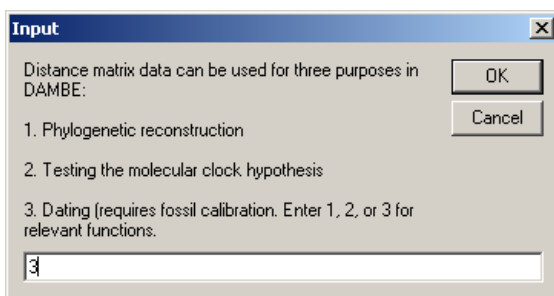


Fig. 12-7. Dialog box for phylogenetic analysis involve distance matrices.

You will be presented with a dialog box similar to that in Fig. 12-3, but without any specification related to the genetic distances. Set the options and click the 'Run' button will generate a dated tree with estimated divergence times.

Dating involving multiple distance matrices is similar. Click 'File|Open other molecular data file', choose 'Distance matrix file', and click the 'OK' button. In the 'Open file' dialog box, choose the 'mtDNAPri3F84SEMmult.dis' file and click the 'Open' button. This file contains 100 distance matrices each generated from bootstrapping the seven sequences in the mtDNAPri.nex file. Proceed exactly the same way as before. You will get a dated tree not only with divergence time estimates, but also with the standard error of each time estimate similar to that shown in Fig. 12-6.

You should try the combination of KumarFig2.dis and KumarFig2.dnd. To appreciate the dramatic difference in evolutionary rate, build a phylogenetic tree from the distance matrix (but enter '1' instead of '3' in Fig. 12-7).

You can copy and paste the tree from DAMBE to PowerPoint slides or any other graphic programs including EXCEL and WORD, either in fixed-resolution bitmapped format or in high-resolution Windows metafile format. The latter should be used for formal presentation or publication. See Appendix 1 Copy trees from DAMBE to PowerPoint slides for details.

I should appreciate your feedback.